

群知能メカニズムを用いた時系列階層型パターン抽出法によるマーケティングデータ分析

Marketing data analysis by hierarchical pattern mining based on swarm intelligence

坪井一晃^{*1} 須賀聖^{*1} 金光博明^{*2} 坂本圭司^{*3} 栗原聡^{*1}
Kazuaki Tsuboi Satoshi Suga Hiroaki Kanemitsu Keiji Sakamoto Satoshi Kurihara

^{*1}電気通信大学
The University of Electro-Communications

^{*2}株式会社トライアルカンパニー
Trial Company, Inc.

^{*3}株式会社トライアルホールディングス
Trial Holdings, Inc.

The behavior patterns in everyday life such as home, office, and commuting, and buying behavior model by day of the week, season, location have hierarchies of various temporal granularity. Generally, in usual hierarchical data analysis, basics hierarchical structure is given in advance. But it is difficult to estimate hierarchical structure beforehand for complex data. Therefore, in this study, we propose algorithm to automatically extract both hierarchical structure and pattern from time series data using swarm intelligent method. We try to extract the patterns from real marketing data.

1. はじめに

創発というキーワードは大変興味深い。創発とは、部品や要素が集合することによって、単独では見られなかった様相が発現することである。生物は創発という現象を上手く活用している。たとえば、アリやハチといった昆虫たちである。それぞれのアリやハチの単体では優れた知能を持つとは言いがたいが、一つの集団として観察すると極めて優れた知能を発揮している。アリを例の一つ具体的に説明する。アリの採餌行動においてアリ単体では餌をみつけ巣に戻るときにはフェロモンという揮発性物質を移動経路に付加することと、餌を探索するにあたりフェロモンが残る経路を好んで探索しやすいという本能的な指針によって行動するのみである。しかし、これを集団として行うことによって、餌場と巣におけるほぼ最短の経路に対して列をなし行進している。このような自然界において単体では単純ながらも集団を形成することで複雑な性質や振る舞いを誘起する様子から着想を得る、群知能の研究が盛んである。

インターネットの急激な進化やコンピュータの高性能化および低価格化を背景に、社会のいたるところにITを活用したシステムが導入されるようになってきている。導入されたシステムを利活用するたび、システムからはデータが生成され続けている。我々が意思を向けずとも稼働しているシステムも存在し、日々多種多様にわたるデータが大量に生成・蓄積されている。近年、この大量に蓄積されるデータを用いることで有用な情報を得られるのではないかと期待が高まっている。このように、蓄積されたデータから有用な知見を得るための技術としてデータマイニングに注目が集まっている。データマイニングにおける重要な技術の一つに、データベースの中から頻出するアイテムの組み合わせを抽出するパターンマイニングがある。パターンマイニングを用いることで、例えばコンビニエンスストアやスーパーマーケットなどの小売店におけるよく同時に購入される商品などを効率的に見つけることができるようになる。さらに、このパターンマイニングを現実世界において度々重視される時系列に着目し発展させた、シーケンシャルパターンマイニング

などの技術が提案されている。時系列性を考慮したパターンを抽出することによって、この商品が購入された次にはこの商品が購入されるはずであるという予測することができるようになる。これらのパターンマイニングでは、解析対象のデータベースはトランザクションと呼ばれる単位で分割されている。したがって、このトランザクションの中という範囲のなかでパターンと考えられるアイテムの組み合わせを抽出している、つまり、データベース中から分析した粒度というものの明確に設定し、それに合わせてトランザクションに分割している。複数の異なる粒度の分析を行うときには、それぞれの設定に応じたトランザクションを生成しその都度分析を行う。

しかし、現実世界を考えたときに、日常における自宅やオフィスなどをつなぐ通勤といった行動様式や小売店における曜日や季節などによる商品の売れ方などの中には、様々な時間的粒度の異なるパターンが含まれていることは明白である。そこで、本研究では、異なる時間的粒度を階層的にとらえ、時系列データから階層構造とパターンの両方を自動的に抽出する手法を提案する。

今回、我々は群知能分野におけるAnt Colony Optimization (ACO) アルゴリズムに着目している。ACO アルゴリズムは自然界におけるアリの群れとしての採餌行動をモデル化したものであり、巡回セールスマン問題を代表とする様々な最適化問題に対して良い性能を示すメタヒューリスティクスであることが知られている。

既存のパターンマイニング技術でも度々問題となる項目として計算量の課題がある。網羅的にパターンを探索する場合、入力データが増えることによって指数関数的に計算量が増加する。そこで、ACO アルゴリズムから着想を得たアルゴリズムを構築することによって、分散処理を行う等の対応を可能とすることで、有限な資源にそった準最適解を得ることに期待できる。また、分析対象が時系列データであるということは、時間経過に伴いパターンが変化することを容易に想定できる。このような問題に対しても、ACO アルゴリズムが有する優れた適応性、頑健性を発揮できるアルゴリズムの構築を試みる。

本稿の構成を次に示す。2章では、パターンマイニングやACO アルゴリズムに関する関連研究について述べる。3章で

連絡先: 坪井一晃, 電気通信大学, 東京都調布市調布ヶ丘 1-5-1, 042-443-5660, tsuboi@ni.is.uec.ac.jp

は、提案手法となる群知能メカニズムによる時系列階層構造を有するパターンの抽出手法について述べる。4章では、提案手法に対してマーケティングデータを用いた実験を行い消費者の購買パターンについて考察する。5章では、まとめと今後の課題について述べる。

2. 関連研究

大量のデータから頻出するアイテムセットの組み合わせをパターンとして抽出する手法がパターンマイニングである。データベース上から頻出するアイテムセットのパターンを抽出するためのアルゴリズムとして Apriori アルゴリズム [1] が有名である。Apriori アルゴリズムは、あるアイテムセットの出現頻度が高いためにはその部分集合のアイテムセットの出現頻度が高い必要があるという考えから、小さいアイテム集合から順にデータベースにおける出現回数を数える幅優先探索を実行する。あるアイテムセットが頻出とならない場合、その上位となるアイテムセットである、そのアイテムセットは包括するより大きいアイテムセットについて探索する必要がないとみなす。このことから解の探索範囲空間を減らしている。しかし、解の候補となるアイテム集合が生成されるたびに出現回数を数えるために全ての入力データベースを調べなければならなくなる。そこで、入力データベースが膨大なものとなったときに、計算時間や計算機のメモリなどに問題が生じる。また、Apriori アルゴリズムでは、アイテムの時系列性を考慮していないといった課題が指摘されている。

そこで、データベース上の時系列性を考慮したアイテムのパターン抽出を行うために Apriori アルゴリズムを拡張した AprioriAll アルゴリズム [2] やパターン抽出にあたって時間制約などの条件をつけることで計算処理の高速化を図った GSP アルゴリズム [3] が提案されている。また、深さ優先型の探索法を採用したシーケンシャルパターンマイニング手法として PrefixSpan アルゴリズム [?] が提案されている。これらのアルゴリズムは計算処理の高速化を図ったものであり、入力データを一様に等しく参照している。しかし、マーケティングなどの現実問題に適用させることを考えると、流行や季節などの影響を受けながら消費者の行動も変化し続けている。このようなパターン自体が変化するような場合には柔軟に対応することが難しい。

そこで、我々は群知能分における ACO アルゴリズムに着目している。ACO アルゴリズムの特徴として、どのような環境に対しても解を生成することができる適応性や、環境の変化に対しても柔軟に解を再探索できる頑健性が挙げられる。このような特徴は現実問題に対して有効であると考えられる。

そもそも、ACO アルゴリズムは自然界におけるアリの採餌行動をモデル化したものであり、最適化問題の巡回セールスマン問題に対する解法として提案された。ACO アルゴリズムにおいて、アリは通過した経路にフェロモンを残すことと、アリはフェロモンの濃度が濃い経路を好んだ経路選択を行うという二つをアリの行動の前提とする。この前提によって、アリが集団として行動するにつれて、最短経路を探索するアリの数が増加しフェロモンの濃度がいっそう濃くなる。一方、フェロモンは揮発性物質であるために時間経過とともに蒸発する。結果的に、アリが通過する頻度が低い経路のフェロモンの濃度は薄くなる。環境に残留するフェロモンが最短経路問題の解となる。結果として、ACO アルゴリズムの適応性や頑健性の高さが特徴となっている。

ACO アルゴリズムをデータマイニングのクラスタリング手

法として応用した研究として Ant Miner+アルゴリズム [4] が提案されている。このアルゴリズムは分類結果がわかりやすい分類器でありながら精度が高く、データが分散した環境にも適するとされる。また、ACO アルゴリズムをパターンマイニングに応用した研究として Tamaki ら [5] の研究がある。Tamaki らは、センサが人の行動を読み取り反応することから、連続した人の行動から隣接するセンサは連続して反応するという前提によって、センサの隣接関係の推定を行っている。ACO アルゴリズムを応用したことで、複数の人による連続してセンサが反応する場合への対応や、センサ自体の故障や移動といった出来事にも柔軟に対応している。

このように、もともとは最適化技術として提案された群知能分野のアルゴリズムであるが、単純な行動ルールに基づいたエージェントの移動と環境に対するフェロモンの付加および蒸発を応用することで、柔軟なシステムの構築が達成できる。本研究においても、ACO アルゴリズムが有する適応性や頑健性に着目し、群知能メカニズムによる時系列階層構造を有するパターンの抽出を試みる。

3. 群知能メカニズムによる時系列階層構造を有するパターンの抽出手法

提案アルゴリズムでは、アイテム集合 $I = \{i_1, i_2, \dots, i_m\}$ で構成させる一つの時系列データ $D_n = \{d_1, d_2, \dots, d_n\}$ (ただし、 $d \subset I$) から、時系列にそって頻出するアイテムの組み合わせをパターンとして順序関係および出現間隔を考慮しながら抽出する。

提案するアルゴリズムでは時系列データから時刻順に投入し、それぞれの入力に合わせて嗜好を有するアリエージェントが探索しフェロモンの付加を行う。また、それぞれの入力に対するアリによるフェロモンの付加が完了するとともに、時間経過となるフェロモンの蒸発を行う。提案するアルゴリズムを擬似コード (アルゴリズム 1) に示す。

Algorithm 1

Input: 時系列データ D_n

Output: フェロモンが付加された仮想マップ M_t

```

 $t := 0$ 
while  $t < n$  do
  アリエージェントの生成・追加
  アリエージェントによる探索および  $M_t$  に対するフェロモンの付加
  フェロモンの蒸発
  探索期間が終了したアリの削除
   $t := t + 1$ 
end while

```

3.1 仮想フェロモンマップ M_t の用意

まず、フェロモンを付加するための環境として仮想マップ M_t を用意する。時刻 t における仮想マップ $M_t = (P, v(P))$ は、パターン集合 $P \ni p$ およびそのパターン p に対するフェロモン量 $v(p)$ を記録した仮想マップである。パターン $p = (F, S, l)$ は、順序関係を考慮した複数のアイテムまたは既出のパターン (F) および、その複数のアイテムまたはパターンの時間間隔 (S)、パターンとしての階階数 (l) からなる。

3.2 アリエージェントの生成

各時刻において、アリエージェントは $ANTS_{max}$ だけ追加生成する。それぞれのアリエージェントはパターンを探索する

にあたって、順序性を有する複数の嗜好 $F \subset \{I, P\}$ とそれぞれの嗜好に対しての好む間隔 S を有する。また、一匹のアリエージェントが探索できる探索期間 r が与えられる。

アリエージェントには2種類あり、1つは最小のアイテムの組み合わせから嗜好を選択するタイプであり、もう1つは生成されたフェロモンマップから嗜好を選択するタイプである、最小のアイテムの組み合わせから嗜好を選択するタイプのアリは、そのアリエージェントが生成された時刻の入力からランダムに第一の嗜好を選択し、与えられる嗜好間のパラメータ分だけ時間経過したときの入力からランダムに第二の嗜好を選択する。

パターンからパターンを階層的に構築するために、フェロモンマップから嗜好を選択するアリエージェントのそれぞれの嗜好 f は、フェロモンマップ M におけるフェロモン量 $v(P)$ の割合を確率としてパターンから選択される。つまり、パターン p が選択される確率 $q(p)$ は式1に従う。

$$q(p) = \frac{v_p}{\sum v(P)} \quad (1)$$

また、アリエージェントに設定する嗜好に対する間隔 S は一様乱数によって定める。アルゴリズムを??に示す。

Algorithm 2 アリエージェントの生成・追加

```

ANT := 0
while ANT < ANTSmax do
  if t < T1 | ANT < ANTθ then
    s := random(1, smax)
    dt から f1 の選択
    dt+s から f2 の選択
  else
    Mt-1 から f1 と f2 の選択
    if f1 を構成する嗜好の2番目のものと f2 を構成する嗜好の1番目のものが同一 then
      s := random(0, smax)
    else
      s := random(1, smax)
    end if
  end if
  f1 探索モードに設定
  ANT := ANT + 1
end while

```

3.3 アリエージェントによる探索及びフェロモンの付加

それぞれのアリエージェントは設定された嗜好および嗜好に対する間隔に基づいて入力された時刻におけるデータに対して探索を行う。アリエージェントが嗜好および嗜好に対する間隔からなるアイテムまたはパターンの組み合わせを発見することができた場合、そのアイテムまたはパターンの組み合わせを一つのパターン p とみなし、そのパターンに対するフェロモン量を1だけ増加させる。なお、発見したパターンが、新規登場であった場合にはパターン番号として順次固有のナンバリングを割り当てる。また、アリエージェントが嗜好およびその間隔から発見したパターンにおいて、アリエージェントが有する嗜好のうち最大の階層数に1加えたものをそのパターンの階層数とする。なお、初期のアイテム集合はそれぞれ階層数を1とする。

Algorithm 3 アリエージェントによる探索

```

ANT := 0
while ANT < ANTSall do
  if f1 探索モードか then
    if Dt に f1 が出現 then
      f2 探索モード (嗜好までのカウント s1) に設定
    end if
  else
    if 嗜好までのカウントが1 then
      if Dt に探索 f が含まれる then
        if fnext があるか then
          fnext 探索モード (嗜好まで snext) に設定
        else
          発見パターンに対してフェロモンの付加
        end if
      else
        f1 探索モードに戻る
      end if
    else
      嗜好まで (のカウント):= -- 1
    end if
  end if
  ANT := ANT + 1
end while

```

3.4 蒸発にともなうフェロモン値の減算

時間経過に伴い、残留するフェロモン量を減少させる。フェロモン量の減少に際して、フェロモン量自体は指数関数的に減少することおよび、パターンのパターンとなっているような深い階層のパターンに関しては出現率が低いと考えられるため、減少するフェロモン量は少ないことを考慮した。

$$v = (\exp(\log(v+1) - 0.00001) - 1) \times (1 - 10^{-l}) \quad (2)$$

4. マーケティングデータを用いた実験

実際のマーケティングデータに対してアルゴリズムを実行し、消費者の購買のパターンについて考察を行う。

4.1 マーケティングデータ

本研究で使用したマーケティングデータは株式会社トライアルホールディングスより提供を受けた、2014年1月1日から2016年12月31日までの3年間の1消費者による商品購買データである。商品購買データにおいては、入力の粒度として商品のカテゴリ名を採用している。また、1日における購買商品を1ステップとし、1095ステップの時系列データとして利用する。実験の対象消費者は3年間のうちに1000日以上購買した経歴がある消費者から、30代女性の消費者を対象とした。

4.2 実験

各ステップ10000のアリエージェントを生成する。アリエージェントに与える嗜好は2つとした。なお、100ステップまでに生成されるすべてのアリエージェントは、生成された時刻において順次パラメータを設定する。また100ステップ以降は100のアリエージェントは生成された時刻において順次パラメータを設定する一方、残りの9900のアリエージェントはフェロモンマップ M_{t-1} から嗜好を選択する。また、2つの嗜好に対する間隔は最大11最小0の一様乱数から選択される。

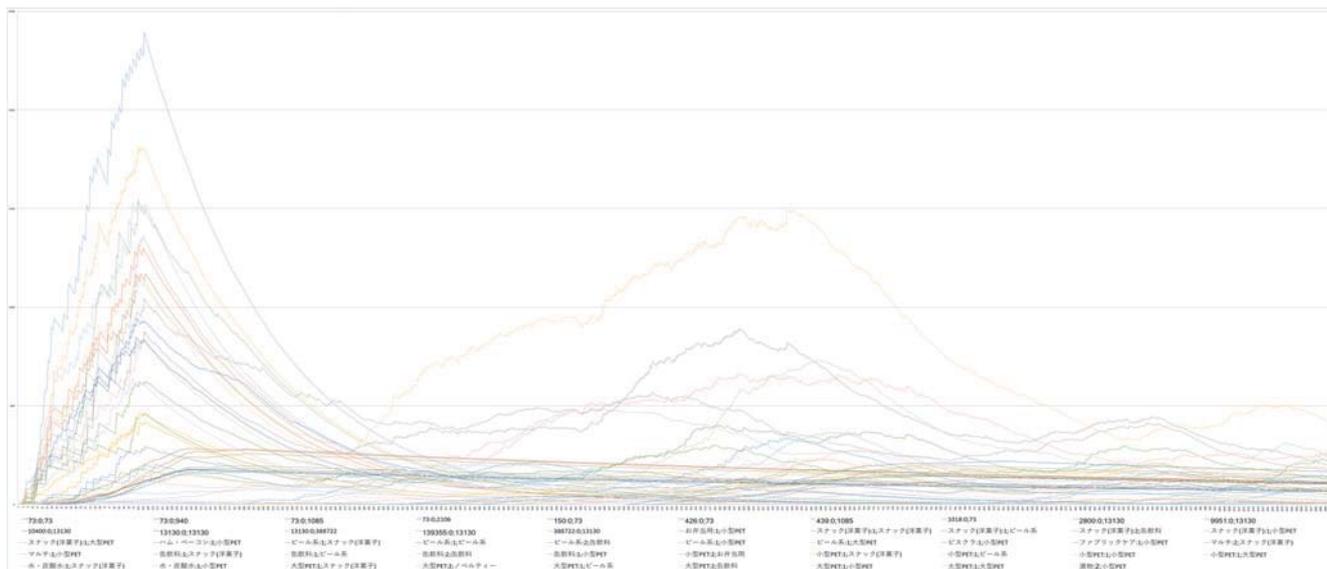


図 1: 30 代女性のパターンのフェロモン量の推移

なお、2つの嗜好に対する間隔において0が選択されるのは、アリエージェントの有する2つの嗜好をみたときに一つ下の階層において重複が許容される場合のみである。

4.3 実験結果

各消費者の商品購買データに対して、提案アルゴリズムを実行した。300, 400, 500, 600, 700, 800, 900, 1000, 1095ステップのそれぞれの時点におけるフェロモン量が多く残る50パターンについて、フェロモン量の推移をそれぞれ図1に示す。ただし、得られたパターンのうち嗜好間の間隔のみが違うものについては、嗜好間の間隔が小さいもののみを取り出した。

4.4 考察

実験の結果、図1より、前半部の100ステップ目を頂点に250ステップまでにかけて、(缶飲料, 2, 缶飲料)に多くのフェロモンが残っているのに対して、それ以降から900ステップまでにかけては(小型PET, 1, 小型PET)に多くのフェロモンが残っていることがわかる。2014年3月あたりまでは缶飲料を隔日で購入していたものが、小型PETボトルを連日購入するようになった。このように、本手法を用いて各消費者の購買行動を分析することで変化する消費者の購買行動に柔軟に抽出できることがわかる。

5. おわりに

本研究では、群知能アルゴリズムに代表されるACOアルゴリズムを基にしたパターン抽出手法を提案するとともに、実際のマーケティングデータを用いて消費者の購買パターンについて抽出する実験を行った。実験の結果、30代女性の隔日で缶飲料を買っていたパターンが、連日の小型ペットボトルを購入するパターンに変化する様子が確認できた。

今後の課題として、アリエージェントに対する適切なパラメータに対する知見の獲得が考えられる。また、現在のアルゴリズムでは、一度アリエージェントにパラメータを設定したらそのパラメータに従い探索し続けるが、各アリエージェントが入力データに対して柔軟に嗜好や嗜好間の間隔などを変化さ

れることによってより適応性を向上させられることも考えられる。

また、実際のマーケティングデータなどの時系列データに対して提案するアルゴリズムを適用させることで、消費者の購買パターンなどの変化する様子を抽出することを目指す。

参考文献

- [1] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." Proc. 20th int. conf. very large data bases, VLDB. Vol. 1215. 1994.
- [2] Agrawal, Rakesh, and Ramakrishnan Srikant. "Mining sequential patterns." Data Engineering, 1995. Proceedings of the Eleventh International Conference on. IEEE, 1995.
- [3] Srikant, Ramakrishnan, and Rakesh Agrawal. "Mining sequential patterns: Generalizations and performance improvements." Springer Berlin Heidelberg, 1996.
- [4] Martens David, et al. "Classification with ant colony optimization" IEEE Transactions on Evolutionary Computation 11.5, pp.651-665, 2007.
- [5] Tamaki, Hiroshi, et al. "Pheromone Approach to the Adaptive Discovery of Sensor-Network Topology." Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 02. IEEE Computer Society, 2008.